A robust optimization approach for the Advanced Scheduling Problem with uncertain surgery duration in Operating Room Planning - an extended analysis

Bernardetta Addis, Giuliana Carello, Elena Tànfani

A robust optimization approach for the Advanced Scheduling Problem with uncertain surgery duration in Operating Room Planning – an extended analysis

Bernardetta Addis[a], Giuliana Carello[b,*], Elena Tànfani[c]

[a] *LORIA (UMR 7503 CNRS), Université de Lorraine, 615 rue du Jardin Botanique, Vandœuvre lès Nancy, France*
[b] *Dipartimento di Elettronica, Informatica e Bioingegneria, Politecnico di Milano, via Ponzio 34, Milano, Italy*
[c] *Dipartimento di Economia, Università di Genova, via Vivaldi 5, Genova, Italy*

Abstract

We consider the Advanced Scheduling Problem (ASP) in the operating room block scheduling, taking into account stochastic patient surgery duration. A surgery waiting list, a set of Operating Room (OR) blocks, and a planning horizon are given. The problem herein addressed is to determine the subset of patients to be scheduled in the considered time horizon and their assignment to the available OR blocks. The problem aims at minimizing a measure of the waiting time of the patients. To this purpose, we introduce a penalty function associated to waiting time, urgency and tardiness of patients.

We propose a robust optimization model to solve the ASP with uncertain surgery durations. The proposed approach does not need to generate a set of scenarios, and guarantees that solutions remain feasible for some variations of the surgery length parameters. We solve the problem on a set of real-based instances to test the behaviour of the proposed model. The solution quality is evaluated with regards to the number of patients operated and their tardiness. Furthermore, assuming lognormal distribution for the surgery times, we use a set of randomly generated scenarios in order to assess the performance of the solutions in terms of OR utilization rate and number of cancelled patients.

Keywords: Operating Room planning, Robust Optimization, uncertain surgery duration, block scheduling.

---

*corresponding author

Email addresses: bernardetta.addis@loria.fr (Bernardetta Addis), giuliana.carello@polimi.it (Giuliana Carello), etanfani@economia.unige.it (Elena Tànfani)

1. Introduction and literature review

In recent years, hospital organizations have been facing a strong pressure to improve the health care delivery processes and to increase their productivity and operational efficiency. In the majority of the hospitals, surgical departments contribute significantly to the total expenditure; besides, they have a great impact on services demands and patient waiting times. The crucial role of surgery departments and their management within hospitals results in an increasing number of research studies aimed at planning Operating Rooms (ORs). Recent literature reviews on operating room planning and scheduling are reported in [1] and [2], where the authors analyze into detail different topics related to the problem settings and summarize significant trends in actual research and possible areas for the future one. Due to the many features that can or cannot be taken into account, several different versions of the OR problem have been considered in literature [3].

OR planning and scheduling problems may be classified according to the scheduling strategy used, i.e. block scheduling, open scheduling, and modified block scheduling. In the block scheduling, each specialty receives a number of OR blocks (usually half-day or full day length) in a given planning period, into which it can arrange its surgical cases [4]. Instead, in the open scheduling, operating rooms are not reserved to a specialty: open scheduling allows surgical cases to be assigned to any operating room available at the convenience of the surgeons or surgical specialties [5]. Modified block scheduling strategy is a mix of the two previous strategies, which can increas the flexibility of the pure block scheduling approach [6].

In this paper we focus on the OR planning and scheduling problem assuming a block scheduling strategy. Within this framework, the problem is usually decomposed into three main phases [7]. Firstly, the number, type and opening hours of the ORs are fixed at a strategic level. Second, the OR capacity is divided among surgical groups or specialties and a cyclic timetable, denoted as *Master Surgical Schedule*, is built on a medium term stand point to account for the tactical assignment of specialties to the OR blocks during the planning horizon. The last phase, referred as *Surgery Process Scheduling*, is divided into two sub-problems: *Advance Scheduling* and *Allocation scheduling* [8, 9]. The Advance Scheduling Problem (ASP) assigns a surgery date and OR to the each scheduled patient, afterwards the allocation scheduling problem determines the sequence of surgeries in each OR block.

We set our analysis at an operational level and we focus our attention on the ASP also known as surgical case assignment, surgery scheduling, surgery admission or surgery loading problem.

Integer and mixed integer linear programming models have been developed for the ASP assuming deterministic surgery times: langragian relaxation approaches [10], branch and price algorithms [11, 12], heuristics [13, 14, 15, 16] and metaeuristics algorithms [17, 18] have been recently proposed.

The OR planning and scheduling problem is further complicated by the inherent variability of the surgical cases durations, which forces the planners to over-conservative scheduling, thus reducing the OR utilization level [19]. Modeling the stochasticity of operating times is a crucial factor in real life planning and scheduling systems, and different assumptions on surgery duration distributions have high impact on the resulting OR overtime and idle time [20].

Fewer papers have been published that propose methods to solve the surgery process scheduling taking into account surgery durations uncertainty. The approaches can be roughly classified into stochastic programming and robust optimization methods. In [21] an advance scheduling problem is considered and uncertainty is managed using a two-stage stochastic model with recourse. The objective function includes the patient waiting times and the OR idle and overtime. The authors compare different heuristics. Furthermore, they also analyze the influence of patient sequencing inside the OR blocks. In [22] a stochastic programming model with recourse is presented. A sample average approximation method to obtain an optimal surgery schedule with the aim of minimizing patient costs and OR overtime costs is used. In [23] a mathematical program considering probabilistic constraints to represent the uncertain duration of surgery procedures is proposed. The proposed model tries to optimize OR utilization without increasing overtime and cancellations. In [24] two models aimed at minimizing the overall OR cost including a fixed cost of opening ORs and a variable cost of overtime are compared. The first is a two-stage stochastic linear model with binary decision variables in the first stage and simple recourse in the second stage. The second is its robust counterpart, in which the objective is to minimize the maximum cost associated with an uncertainty set for surgery durations. They show that the robust method is much faster than solving the stochastic recourse model, and has the benefit of limiting the worst-case outcome of the recourse problem. In [25] different heuristics for the robust surgery loading problem are proposed, with the aim of maximizing the utilization of operating theatre and minimizing the overtime risk by introducing planned slack times. In [26] a two-level framework is proposed. In the first level, a MIP model finds a deterministic solution for the OR planning problem. In the second level, the variability of surgery duration is taken into account by means of individual chance constraints for each OR block and a robust solution is achieved by iteratively adding safety slacks to the first level deterministic model solutions.

Simulation based approaches are also proposed in literature. Some authors use simulation to compare different scheduling strategies and test the solution robustness against the randomness of surgery duration [27, 7, 28]. Although the majority of the authors restricts their analysis to the evaluation of alternative scenarios, advanced simulation-based optimization approaches have been proposed combining simulation with other solution techniques [29, 30].

In this paper, we propose a robust optimization approach to select and assign a set of waiting patients to OR blocks in a given planning horizon, assuming that patient operating times are random parameters. The aim is to minimize a measure of the total waiting time. Based on the seminal work [31], we propose a penalty function that takes into account waiting time, urgency and tardiness of all patients to be scheduled. The robustness of solutions is achieved by applying the approach proposed in [32], which allows to exploit the potentialities of a linear programming model without the necessity of generating scenarios. This approach does not require to know the probability density functions of the surgery duration. It requires only limited information and few general assumptions which is a realistic limitation in many real-based application. We propose the formulations of the deterministic and robust version of the problem. The models are compared over a set of real life based instances to evaluate their behavior in terms of computational effort and solution quality. The solution quality is also evaluated with regards to the number of patients operated and their waiting time and tardiness. Moreover, assuming lognormal distributions for the surgery times, a set of randomly generated scenarios is used in order to compare the proposed solutions in terms of OR utilization rate and number of cancelled patients. The impact of introducing overtime in the model formulation is evaluated and a sensitivity analysis on the choice of the key parameters is performed.

The remainder of the paper is organized as follows: in Section 2 we introduce the problem under investigation and the deterministic and robust formulation are presented. In Section 3, the results on a set of real-based randomly generated instances are reported and compared. Finally, in Section 4 conclusions and future research directions are given.

2. Problem description and models

In the ASP a set of elective patients I is given to be scheduled in a planning horizon. Let D be the length in days of the planning horizon. We assume a block scheduling approach and focus on a single surgical specialty, but the approach can be easily adapted to take into account more than one specialty. A set J of OR blocks assigned to the specialty and their schedule during a week are given. Each block is described by an operating room and a week day. The planning horizon is then represented by a sequence of repetitions of the same group of blocks in a set of weeks K. The available total time of a time block j in week k, i.e. the OR block length, is denoted as γ_{jk}.

The patients in the set I belong to a waiting list, where patients are registered at the moment they arrive at the service. For each patient i, let w_i denote the number of days which the patient has already spent in the waiting list at the beginning of the planning horizon. Moreover, a maximum waiting time l_i and a corresponding urgency parameter u_i are given for each patient i. If the patient has spent w_i days in the waiting list, he/she must receive surgery before a due date $dd_i = l_i - w_i$, otherwise he/she is considered tardy. According to the block weekly based pattern, if a patient is scheduled in block $j \in J$ and week $k \in K$, he/she waits a total number of days $d_{jk} = 7(k-1) + j$. The surgery time $\tilde{t}_i$ for each patient i is consider to follow a given probability distribution.

The Stochastic Advanced Scheduling (SAS) problem can be defined: select a subset of patients to be operated on in the considered planning horizon and assign them to weeks and OR blocks, while guaranteeing that the capacity of each block is not exceeded. The objective function aims at minimizing an overall penalty due to delay in serving the patients. As proposed in [15] it takes into account both the urgency and waiting time of scheduled and not scheduled patients. Besides, a penalty for due date violation and patient tardiness is also considered ([31]).

The problem can be formulated using the following set of binary variables x_{ij}^k, such that $x_{ij}^k = 1$ if patient i is assigned to block j in week $k \in K$, and zero otherwise. The objective function is formulated as follows:

$$\min \quad \sum_{i\in I}\left\{\sum_{j\in J}\sum_{k\in K}\left[d_{jk}+(w_i+d_{jk}-l_i)^+\right]u_i x_{ij}^k \right. \tag{1}$$
$$\left. +\left[(w_i+D+1)+(w_i+D+1-l_i)^+\right]u_i\left(1-\sum_{j\in J}\sum_{k\in K}x_{ij}^k\right)\right\},$$

where $(w_i+d_{jk}-l_i)^+=\max\{w_i+d_{jk}-l_i,0\}$ is the patient tardiness, that is the number of days waited after the due date. The first term represents the penalty for the scheduled patients. For each scheduled patient i the penalty is composed by two parts: the number of days d_{jk} spent before receiving surgery in the planning horizon and the tardiness $(w_i+d-l_i)^+$ of the patient. The term is weighted by the patient urgency parameter u_i, in order to give priority to the most urgent patients. The second term is associated with the penalty of the unscheduled patients. It is the sum of the tardiness and the overall days spent waiting for surgery before and after the beginning of the planning horizon, while for the scheduled patients, the waiting days term consider also the days after the beginning of the planning horizon. As real tardiness and waiting days cannot be computed for unscheduled patients (we do not know when there will be scheduled), we use a lower bound to take them into account, which is calculated assuming that all the remaining patients are scheduled the first day after the planning horizon $(D+1)$. Also for the unscheduled patients the waiting time and the tardiness are weighted by the urgency parameter u_i.

The set of constraints is the following:

$$\sum_{j\in J}\sum_{k\in K}x_{ij}^k\le 1 \qquad \forall i\in I \tag{2}$$
$$\sum_{i\in I}\tilde{t}_i x_{ij}^k\le\gamma_{jk} \qquad \forall j\in J,\quad\forall k\in K \tag{3}$$

Constraints (2) ensure that each patient is operated at most once. Constraints (3) are the stochastic capacity constraints for each block forcing the total time in block j of week k to be lesser than or equal to the maximum available time γ_{jk}.

The Deterministic Advanced Scheduling (DAS) model is obtained from the SAS model using for each patient i a deterministic surgery time $\bar{t}_i$. Constraints (3) are replaced by

$$\sum_{i\in I}\bar{t}_i x_{ij}^k\le\gamma_{jk} \qquad \forall j\in J,\quad\forall k\in K \tag{4}$$

Beside we propose a robust counterpart, the Robust Advanced Scheduling (RAS) model to deal with uncertainty, which is based on the robust optimization approach proposed in [32].

According to [32], a random variable is assumed to vary in a given interval. Uncertainty is dealt with so as to guarantee than any solution is feasible if, for each constraint involving uncertain parameters, at most Γ of them assume the maximum value in the interval and all the others assume the central one. In our case the uncertain parameters are the uncertain surgery times $\tilde{t}_i$. We consider the interval $\bar{t}-\hat{t}, \bar{t}+\hat{t}$, where the central value of the interval is denoted as $\bar{t}$ and the maximum value we want to protect is equal to $\bar{t}_i + \hat{t}_i$. In the computational results we will propose and discuss possible choices for $\hat{t}$. Then, for each block j, and week k, a subset S_{jk} of patients, who require their maximum surgery time, such that $|S_{jk}| = \Gamma$, is chosen among the patients assigned to the block in the given week. Among all the possible subsets, the one having the worst impact on the capacity constraint is selected, and the solution is guaranteed to be feasible even with respect to this subset:

$$\sum_{i \in I} \bar{t}_i x_{ij}^k + \max_{S_{jk} \subset I : |S_{jk}| = \Gamma} \left\{ \sum_{i \in S_{jk}} \hat{t}_i x_{ij}^k \right\} \leq \gamma_{jk} \qquad \forall j \in J, \quad \forall k \in K \tag{5}$$

The value $\max_{S_{jk} \subset I : |S_{jk}| = \Gamma} \left\{ \sum_{i \in S_{jk}} \hat{t}_i x_{ij}^k \right\}$ can be computed for each block j and each week k solving the following Linear Programming problem:

$$\beta^{jk} = \max \left(\sum_{i \in I} \hat{t}_i x_{ij}^k \right) z_i \tag{6}$$

$$\sum_{i \in I} z_i \leq \Gamma \tag{7}$$

$$z_i \leq 1 \qquad \forall i \in I \tag{8}$$

$$z_i \geq 0 \qquad \forall i \in I \tag{9}$$

Let denote with ζ^{jk} the dual variables associated to constraints (7) and with π_i^{jk} the dual variables associated to constraints (8). The dual of (β^{jk}) can be formulated as follows:

$$\min \quad \Gamma \zeta^{jk} + \sum_{i \in I} \pi_i^{jk} \tag{10}$$

$$\zeta^{jk} + \pi_i^{jk} \geq \hat{t}_i x_{ij}^k \qquad \forall i \in I \tag{11}$$

$$\zeta^{jk}, \pi_i^{jk} \geq 0 \tag{12}$$

The optimal values of the objective functions (6) and (10) coincide. Thus, constraints (3) can be linearized, by replacing them with:

$$\sum_{i \in I} \bar{t}_i x_{ij}^k + \Gamma \zeta^{jk} + \sum_{i \in I} \pi_i^{jk} \leq \gamma_{jk} \qquad \forall j \in J \quad \forall k \in K \tag{13}$$

$$\zeta^{jk} + \pi_i^{jk} \geq \hat{t}_i x_{ij}^k \qquad \forall j \in J \quad \forall i \in I \tag{14}$$

$$\zeta^{jk}, \pi_i^{jk} \geq 0 \tag{15}$$

The resulting RAS model formulation is as follows:

$$\min \sum_{i\in I}\left\{\sum_{j\in J}\sum_{k\in K}[d_{jk}+(w_i+d_{jk}-l_i)^+]u_i x_{ij}^k \right.$$

$$\left. +[(w_i+D+1)+(w_i+D+1-l_i)^+]u_i\left(1-\sum_{j\in J}\sum_{k\in K}x_{ij}^k\right)\right\} \tag{16}$$

$$\sum_{j\in J}\sum_{k\in K}x_{ij}^k \leq 1 \qquad \forall i\in I \tag{17}$$

$$\sum_{i\in I}\bar{t}_i x_{ij}^k + \Gamma\zeta^{jk} + \sum_{i\in I}\pi_i^{jk} \leq \gamma_{jk} \qquad \forall j\in J, \forall k\in K \tag{18}$$

$$\zeta^{jk}+\pi_i^{jk} \geq \hat{t}_i x_{ij}^k \qquad \forall j\in J, \forall i\in I \tag{19}$$

$$\zeta^{jk}, \pi_i^{jk} \geq 0 \qquad \forall j\in J, \forall k\in K, \forall i\in I \tag{20}$$

$$x_{ij}^k \in \{0,1\} \qquad \forall j\in J, \forall k\in K, \forall i\in I \tag{21}$$

3. Experimental tests

The deterministic and robust formulations have been tested in order to evaluate the applicability of the proposed approach both in terms of computational effort and quality of the obtained solutions.

The models have been tested on a set of instances derived from real life data. First, the obtained solutions are compared with respect to the objective function value and the number of operated patients, in order to evaluate the impact of different values of Γ, and thus different levels of required robustness. Then, the obtained assignments of patients to OR blocks are evaluated on a set of 100 randomly generated scenarios. The aim is to study the behavior of the proposed solutions in terms of utilization rate and number of cancelled patients. Further, a sensitivity analysis is performed to evaluate the impact of different choices of the maximum deviation of surgery times. Tests on larger instances are run to evaluate the computational effort required in solving the models.

Finally, the possibility of allowing overtime is considered. Both formulations are modified to take into account block overtime. The obtained solutions are evaluated on the 100 scenarios in terms of utilization rate and cancelled patients.

Instances and scenarios generation are described in Section 3.1. Models results are reported and analysed in Section 3.2, while the overtime impact in discussed in Section 3.3.

3.1. Instances and scenarios generation

The instances are generated from real data based waiting lists partially derived from [26]. Each waiting list represents a collection of patients who wait for surgery and should be scheduled.

In the following we refer to an already validated prioritisation system based on five urgency classes [33]. Each urgency class is associated to a maximum waiting time expressed in days, that is the maximum number of days that a patient can wait before surgery without deteriorating his/her clinical conditions. The maximum waiting times are set to 8, 30, 60, 180 and 360 days, respectively. The maximum waiting time contributes in defining the urgency coefficient, which represents the speed at which the clinical need is assumed to increase with time. In particular, for each class the urgency coefficient is computed as the ratio between the maximum waiting time of the least urgent class and its own maximum waiting time. The corresponding urgency coefficients of the five considered classes are 45, 12, 6, 2, 1, respectively.

For each waiting list, we generated eight instances by assigning different surgery times to patients. Real data surgery time were derived from [25]. We used data of eight different surgery time lists. Each list is composed by different types of surgery. Each surgery type is described by an average surgery time, a standard deviation and the percentage of this type over the total number of surgeries in the list. According to these percentages, an average surgery time ($\bar{t}_i$) and a standard deviation (σ_i) is randomly assigned to each patient i. We recall that the maximum value we want to protect from ($\hat{t}$) must be determined to apply the robust approach. The maximum deviation is assumed to be proportional to the standard deviation $\hat{t} = \alpha\sigma_i$.

Each instance represents the combination of a waiting list and a surgery list. Each instance is named n-s, where n is the the number of patients ($|I|$), and s is the list index used for the surgery times generation.

We consider a 7 days time horizon, corresponding to one week.

For the first series of computational results, we considered two sets of instances: one with 20 patients and two OR blocks per week, scheduled on Monday and Wednesday, the other with 40 patients and three blocks per week, on Monday, Wednesday and Friday. Each block j is assumed to have a capacity of 6 hours.

For each instance we generated 100 different random realizations. In each realization, for each patient i the surgery time $\tilde{t}_i$ is randomly generated using a lognormal distribution with average surgery time $\hat{t}_i$ and standard deviation σ_i. Lognormal distribution is widely applied in describing stochastic surgery times. In [34] it is showed that plots of surgical times reveal a truncation on the left side and a long tail in the right hand side suggesting a lognormal distribution. As well, in [35] a lognormal distribution with non-zero minimum parameter is suggested to represent procedure times. In [36] normal and lognormal models to represent surgical procedure times are compared, and the use of the lognormal distribution for predicting these times is recommended. Furthermore, in [37] the lognormal distribution to estimate prediction bounds to support managerial decision making on the day of surgery is used.

		DAS	$\Gamma = 1$	$\Gamma = 2$	$\Gamma = 3$	$\Gamma = 4$	$\Gamma = 5$	$\Gamma = 6$	$\Gamma = 7$	$\Gamma = 8$
20 patients	average	0.02	0.05	0.10	0.21	0.11	0.19	0.18	0.17	0.09
	max	0.03	0.08	0.20	1.14	0.34	0.73	0.59	0.56	0.37
40 patients	average	0.15	6.80	2.06	11.38	102.73	107.67	250.81	73.85	172.00
	max	0.82	52.59	10.32	42.37	613.86	645.20	1678.96	521.78	1334.47

Table 1: Computational times in seconds

To avoid too short surgery times, we truncate the lognormal distribution at a minimum value equal to $\max(\bar{t}_i - \sigma_i, 30)$. If r_i is the random generated number following the lognormal distribution, the surgery time assigned to patient i will be: $\tilde{t}_i = \max(r_i, \bar{t}_i - \sigma_i, 30)$.

For the larger instances tests, we considered three waiting lists, with 80, 120 and 140, respectively. The larger instances are generated with the same procedure applied for the smaller ones. For each waiting list, two sets of instances are considered, one with three blocks per week, and one with five blocks per week.

3.2. Models results

The deterministic and robust models are tested on the set of instances described in Section 3.1. The models have been implemented with AMPL and solved with CPLEX 12.2.0.0 on a Intel Xeon CPU E5335 (2 quad core cpus at 2GH). We set a 2 hours time limit and a $1.e-3$ accetable relative gap. Eight values of Γ have been considered, from 1 to 8. Results on the objective function and number of operated patients are discussed in Section 3.2.1. The behavior of the obtained solutions over the random scenarios is analysed in Section 3.2.2. We set $\alpha = 1$ for these first tests, i.e. the maximum deviation is assumed to be equal to the standard deviation; then we performed a sensitivity analysis on α parameter in Section 3.2.3. Finally, tests on larger instances are reported in Section 3.2.4 to evaluate the models with respect to computational time and scalability.

3.2.1. Objective function and operated patients

In Table 1 the average and maximum computational times are given, for DAS and RAS models and for each value of Γ. All the considered instances have been solved to optimality within the time limit. Solving the deterministic model requires few seconds, while the computational time may significantly vary for the robust counterpart. Computational time is higher for higher values of Γ, with a peak corresponding to $\Gamma = 5$ or $\Gamma = 6$. However, the required CPU time is never above one hour and a half.

The trend of the objective function is reported in Figure 1 and 2, for the 20 and 40 instances, respectively. In particular, the figures represent the average increase, w.r.t. the non robust case, of the waiting time of each patient in number of urgency weighted days. The value increases for larger values of Γ: indeed a more robust solution is required, a larger subset of patients requires the maximum surgery time and therefore less patients are scheduled per day. However the average delay, and in general the objective function, is constant after a value $\underline{\Gamma}$ which varies for the different instances considered. This is due

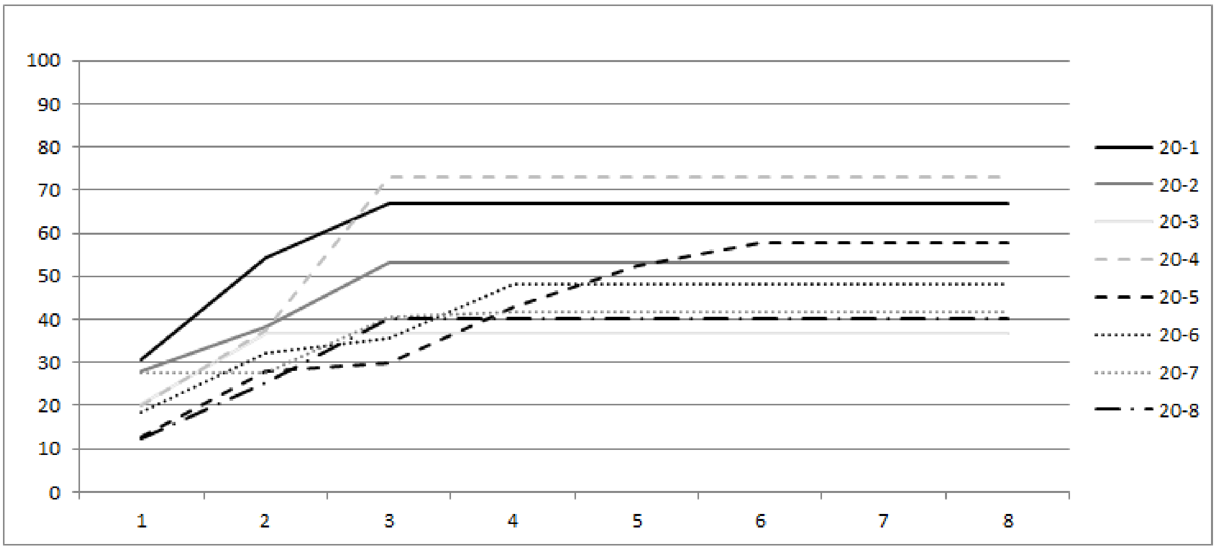

Figure 1: Average per patient delay for 20 patients instances

to the fact that when the value of Γ is equal to, or greater than, the maximum number of patients which can be scheduled in one block, all such patients are assumed to ask for their maximum surgery time. In the following tables, the values after $\underline{\Gamma}$ are replaced by "-".

In Table 2 the number of operated patients is given for each instance and value of Γ. In the first column (max) an upper bound of the number of patients who can be operated is also reported. Such value is computed by maximizing the number of operated patients without taking into account patient penalties.

Results show that if robustness is not required (DAS) the number of operated patients is close to the maximum possible. The number of operated patients usually decreases with the increasing values of Γ. However, for low values of Γ the value is still very close to the bound. Even for higher values of Γ it is less than half the bound only in few instances, while in general it is more than 50%. Note that the set of operated patients is different for different values of Γ, and, in general, the set of operated patients for $\Gamma = m$ is not a subset of those operated for $\Gamma = m - 1$.

Table 3 reports about those patients whose deadline is exceeded (tardy patients). The number of patients who receive their surgery in the planning horizon after due date is reported in the "op" columns, while the number of patients who are tardy, but will not be scheduled in the planning horizon is given in the "nop" columns. As the objective function takes into account the weighted waiting times, rather than the number of tardy patients, the number may increase or decrease with the increasing value of Γ parameter. In general, the variation is not dramatical in most of the instances. Furthermore, the positive impact of requiring robustness will be highlighted by tests on realizations.

3.2.2. Utilization rate and cancelled patients

The behavior of the obtained solutions on a set of 100 randomly generated scenarios is described in Table 4 and in Table 5. In particular, in Table 4 the operating room utilization

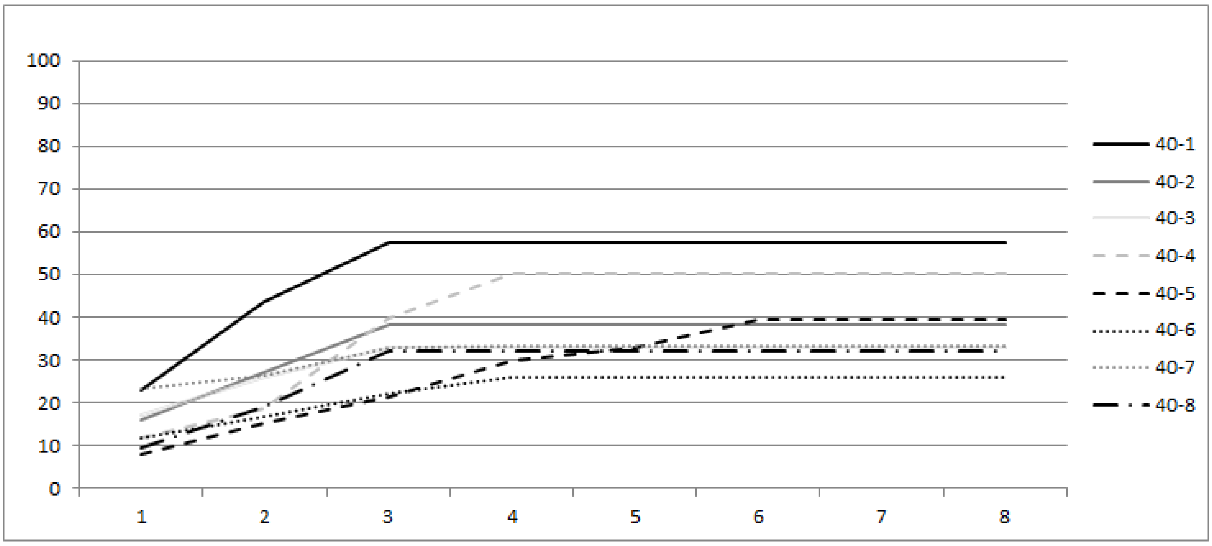

Figure 2: Average per patient delay for 40 patients instances

Instance	max	DAS	$\Gamma = 1$	$\Gamma = 2$	$\Gamma = 3$	$\Gamma = 4$	$\Gamma = 5$	$\Gamma = 6$	$\Gamma = 7$	$\Gamma = 8$
20-1	8	8	6	6	4	-	-	-	-	-
20-2	9	8	8	7	6	-	-	-	-	-
20-3	7	7	6	5	-	-	-	-	-	-
20-4	10	9	8	8	6	-	-	-	-	-
20-5	12	11	10	9	10	9	9	8	-	-
20-6	8	7	7	6	6	5	-	-	-	-
20-7	8	7	7	7	6	-	-	-	-	-
20-8	8	6	7	6	5	-	-	-	-	-
40-1	14	12	9	7	5	-	-	-	-	-
40-2	15	13	11	12	10	-	-	-	-	-
40-3	12	10	9	8	7	-	-	-	-	-
40-4	15	14	13	12	10	9	-	-	-	-
40-5	20	18	17	15	15	13	13	12	-	-
40-6	14	10	9	10	9	8	-	-	-	-
40-7	13	10	11	8	-	-	-	-	-	-
40-8	13	10	9	9	7	-	-	-	-	-

Table 2: Number of operated patients

Instance	DAS		Γ =1		Γ =2		Γ =3		Γ =4		Γ =5		Γ =6		Γ =7		Γ =8	
	op	nop	op	nop	op	nop	op	nop	op	nop	op	nop	op	nop	op	nop	op	nop
20-1	4	5	4	6	4	6	5	6	5	6	5	6	5	6	5	6	5	6
20-2	6	3	6	4	5	4	4	5	4	5	5	5	5	5	5	5	5	5
20-3	6	4	5	5	5	5	5	5	4	6	4	6	4	6	4	6	4	6
20-4	9	0	7	1	8	2	7	3	5	4	6	4	6	4	6	4	6	4
20-5	8	1	7	2	7	2	7	2	7	2	7	2	5	3	7	2	7	2
20-6	4	4	5	5	4	5	5	5	4	5	4	5	4	5	4	5	4	5
20-7	5	4	4	5	5	5	5	6	4	5	4	5	4	5	4	5	4	5
20-8	5	5	4	5	5	6	4	6	5	6	5	6	5	6	5	6	5	6
40-1	8	4	8	5	7	6	6	7	6	7	6	7	6	7	6	7	6	7
40-2	11	1	11	2	10	3	8	4	10	3	10	3	10	3	10	3	10	3
40-3	10	3	9	4	9	4	8	5	8	5	8	5	8	5	8	5	8	5
40-4	11	0	12	0	11	1	11	2	10	2	10	3	10	3	10	3	10	3
40-5	10	1	11	1	11	1	11	1	10	1	10	1	10	1	11	1	11	1
40-6	9	3	8	4	8	5	9	4	8	5	8	5	8	5	7	5	8	5
40-7	10	3	8	5	8	5	8	5	8	5	8	5	8	5	8	5	8	5
40-8	9	4	9	4	8	5	7	6	8	5	7	6	7	6	7	6	7	6

Table 3: Number of patients operated (op) and still waiting (nop) after their due date

rate is given, while in Table 5 the average minimum number of cancelled patients for each block is reported. Results show that the operating rooms are well exploited if no robustness is required (DAS). The utilization rate is between 90% and 100% for DAS case, while it decreases when the value of Γ increases, since longer surgery times are considered for at least a subset of patients. The rate may fall to about 70% for most of the instances, but it is always above 50%. On the other hand, with small values of Γ the number of cancelled patients is significant. It rises up to more than 0.4 in five instances if no robustness is required, and it reaches 0.65 for instance 40-8. In the robust solutions, the number of cancelled patients decreases when the value of Γ increases. The selected assignment is almost completely respected for $\Gamma \geq 4$. For example, let us consider instance 20-1. The deterministic solution produces an average number of cancelled patients per block equal to 0.47. As there are two blocks in a week, about one patient is cancelled per each week. On the other side, by setting $\Gamma = 3$, the number of cancelled patients is about 3 every 50 weeks (less then 4 cancellations in one year).

By properly tuning the value of Γ a tradeoff between the utilization rate and the number of cancelled patients can be obtained. In fact, from the hospital management point of view, smaller values of Γ are preferable, as they guarantee a higher utilization rate. However, such values impact on the solution robustness, as it is shown by the higher number of cancelled patients. Cancelled patients must be reschedule in a longer term period planning. If cancelled patients are too frequent, the planning is disrupted and must be re-optimized. From the perceived quality of service point of view, instead, higher values of Γ are better as they guarantee that the OR schedule is respected and no patients must be delayed from the plan (and rescheduled in further periods). Besides, it is worth noting that an utilization rate below 100% means that there is some operating room capacity not utilized. Such available OR time, rather than being a loss for the system, could allow to manage emergency cases and/or cancelled patients to be rescheduled, without changing the planned OR schedule.

3.2.3. Maximum deviation sensitivity

We already pointed out that for the robust approach we used, a maximum value $\hat{t}$ for uncertainty parameter has to be chosen. We assume $\hat{t} = \alpha\sigma_i$. In the first computational experiments we set $\alpha = 1$. We now tested the impact of different values of parameter α, that is to say, of different choices of the maximum deviation. We run our model with $\alpha = 0.5$, $\alpha = 1$, and $\alpha = 2$. The behavior of the average delay per patient is reported in Figure 3. Results show that the impact of α and Γ are combined, as they both have an impact on the overall required surgery time. For a given value of Γ, increasing the value of α increases the average delay. The trend is similar for 20 and 40 patients instances. The value of $\underline{\Gamma}$ in general decreases with the increasing value of α.

The obtained solutions are then tested on the scenarios. Results on the utilization rate and on the number of cancelled patients are shown in Figure 4 and 5. As for the average delay, the impact of the two parameters is combined. For a given value of Γ, the utilization rate decreases with the increasing value of α. Increasing the value of α reduces significantly the number of cancelled patients. For $\alpha = 2$ the number of cancelled patients is equal zero

Instance	DAS	Γ =1	Γ =2	$\Gamma = 3$	$\Gamma = 4$	$\Gamma = 5$	$\Gamma = 6$	$\Gamma = 7$	$\Gamma = 8$
20-1	101.72	69.77	62.06	54.54	-	-	-	-	-
20-2	99.17	83.53	77.03	69.73	69.73	69.73	70.08	-	-
20-3	97.74	83.56	72.16	-	-	-	-	-	-
20-4	94.30	84.93	80.11	63.02	63.02	63.22	63.02	-	-
20-5	95.81	89.20	82.94	82.82	76.25	74.36	70.28	69.99	70.28
20-6	99.86	87.62	78.86	77.04	68.91	-	-	-	-
20-7	92.53	79.84	79.84	68.51	69.28	69.28	69.44	69.28	69.44
20-8	91.42	80.82	70.95	61.30	-	-	-	-	-
40-1	102.11	73.82	64.80	49.10	-	-	-	-	-
40-2	101.11	84.85	80.19	70.75	70.62	70.67	70.67	70.67	70.70
40-3	97.83	81.41	73.53	69.86	-	-	-	-	-
40-4	96.34	87.38	81.47	67.87	62.17	62.10	61.92	61.93	62.11
40-5	99.94	92.77	85.18	83.18	75.70	74.59	70.13	70.11	70.08
40-6	94.71	86.47	81.04	75.56	76.36	75.92	75.92	76.36	-
40-7	91.39	86.28	69.25	62.69	62.74	-	-	-	-
40-8	92.53	79.84	79.84	68.51	69.28	69.28	69.44	69.28	69.44

Table 4: Utilization rate in percentage

Instance	DAS	Γ =1	Γ =2	$\Gamma = 3$	$\Gamma = 4$	$\Gamma = 5$	$\Gamma = 6$	$\Gamma = 7$	$\Gamma = 8$
20-1	0.47	0.13	0.13	0.03	-	-	-	-	-
20-2	0.32	0.19	0.07	0.06	0.06	0.06	0.03	-	-
20-3	0.19	0.02	-	-	-	-	-	-	-
20-4	0.25	0.07	0.08	0.00	0.00	0.01	0.00	-	-
20-5	0.21	0.12	0.05	0.03	0.03	0.02	0.00	0.02	0.00
20-6	0.15	0.11	0.05	0.02	0.00	-	-	-	-
20-7	0.25	0.05	0.05	0.02	0.00	-	-	-	-
20-8	0.29	0.10	0.04	0.01	-	-	-	-	-
40-1	0.26	0.07	0.08	0.01	0.00	0.00	0.01	-	-
40-2	0.41	0.16	0.08	0.04	0.05	0.05	0.05	0.06	0.05
40-3	0.17	0.05	0.03	0.01	-	-	-	-	-
40-4	0.41	0.26	0.12	0.04	0.00	-	-	-	-
40-5	0.43	0.16	0.08	0.06	0.00	0.00	0.00	0.00	0.01
40-6	0.26	0.08	0.03	0.00	0.02	-	-	-	-
40-7	0.25	0.07	0.03	0.00	0.01	-	-	-	-
40-8	0.65	0.31	0.12	0.06	-	-	-	-	-

Table 5: Average number of cancelled patients per block

for almost all the instances even for $\Gamma = 2$.

The average per patient delay, the utilization rate and the number of cancelled patients for a particular instance, namely instance 20-5, are reported in Figure 6, 7 and 8, respectively. The analysis allows to highlight the impact of parameter values and the way in which they are combined. Let us considered for instance the average delay (Figure 6) for $\Gamma = 1$ and $\Gamma = 2$. A delay of about 30 is obtained either for $\Gamma = 1$ and $\alpha = 2$, or for $\Gamma = 2$ and $\alpha = 1$. A value slightly below 60 is obtained either for $\Gamma = 2$ and $\alpha = 2$, or for $\Gamma = 6$ and $\alpha = 1$. Thus, the two value impacts are combined, but a high value of α produces behaviors which cannot be obtained for lower values by increasing Γ.

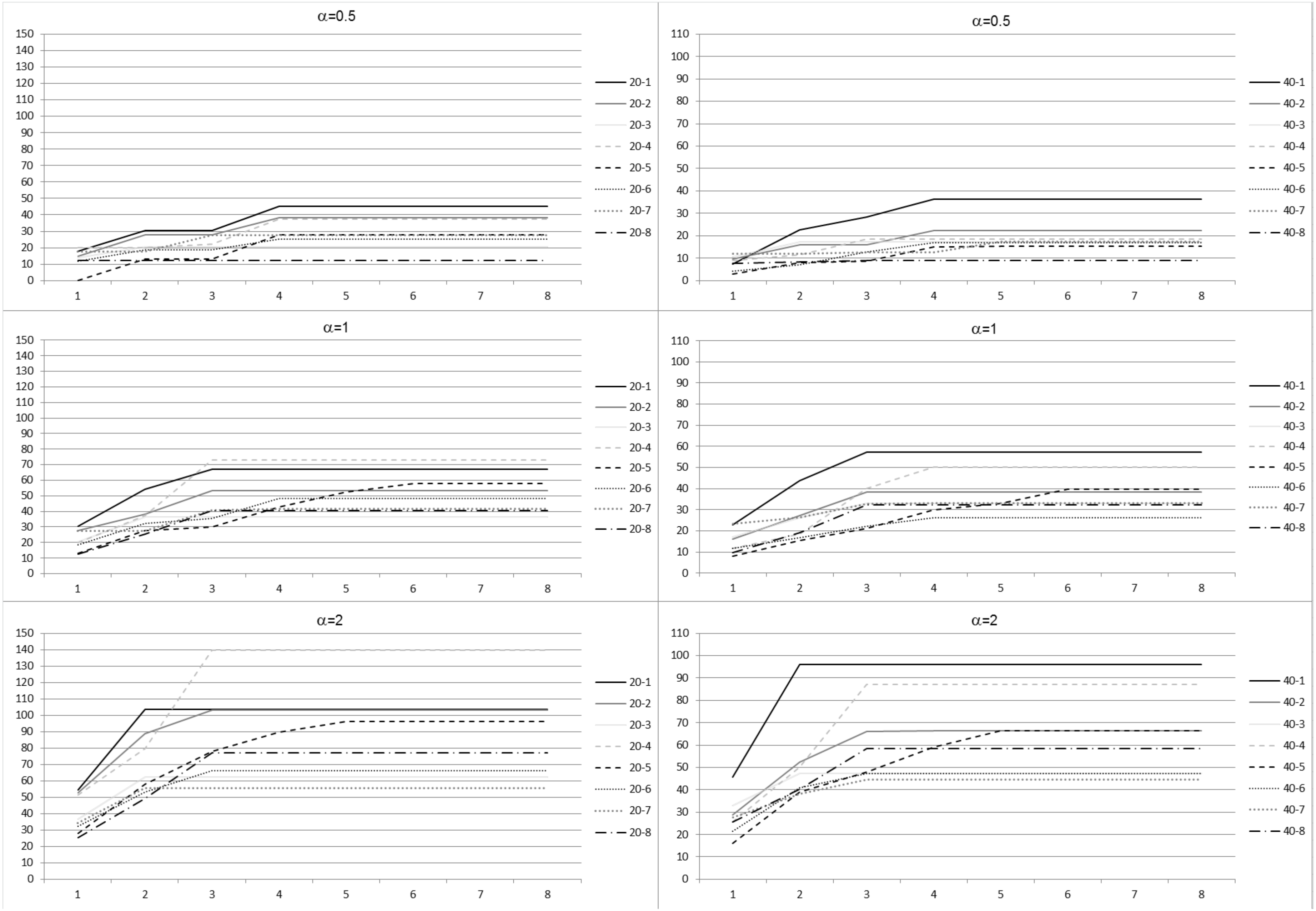

Figure 3: Average per patient delay for different values of α

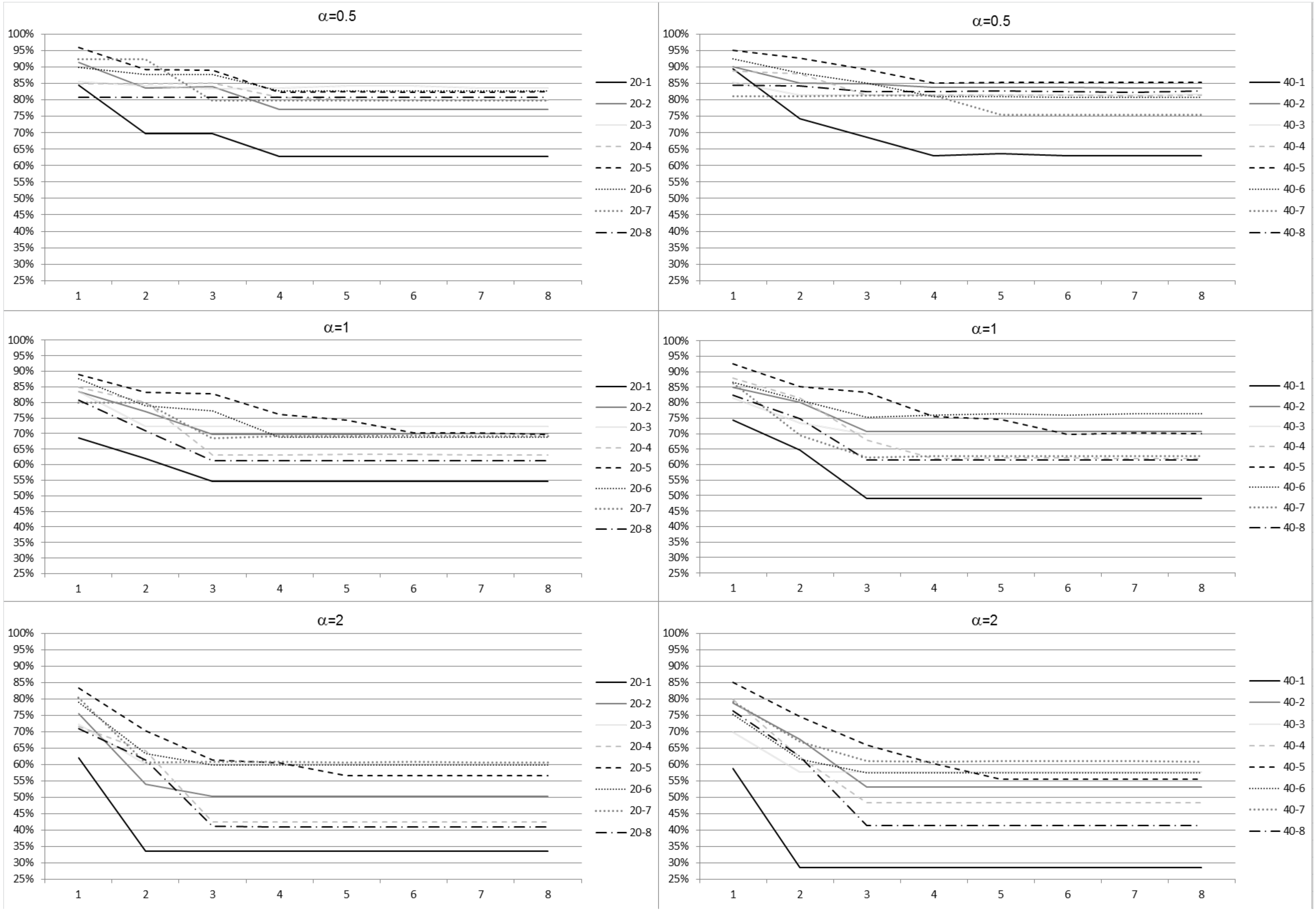

Figure 4: Utilization rate for different values of α

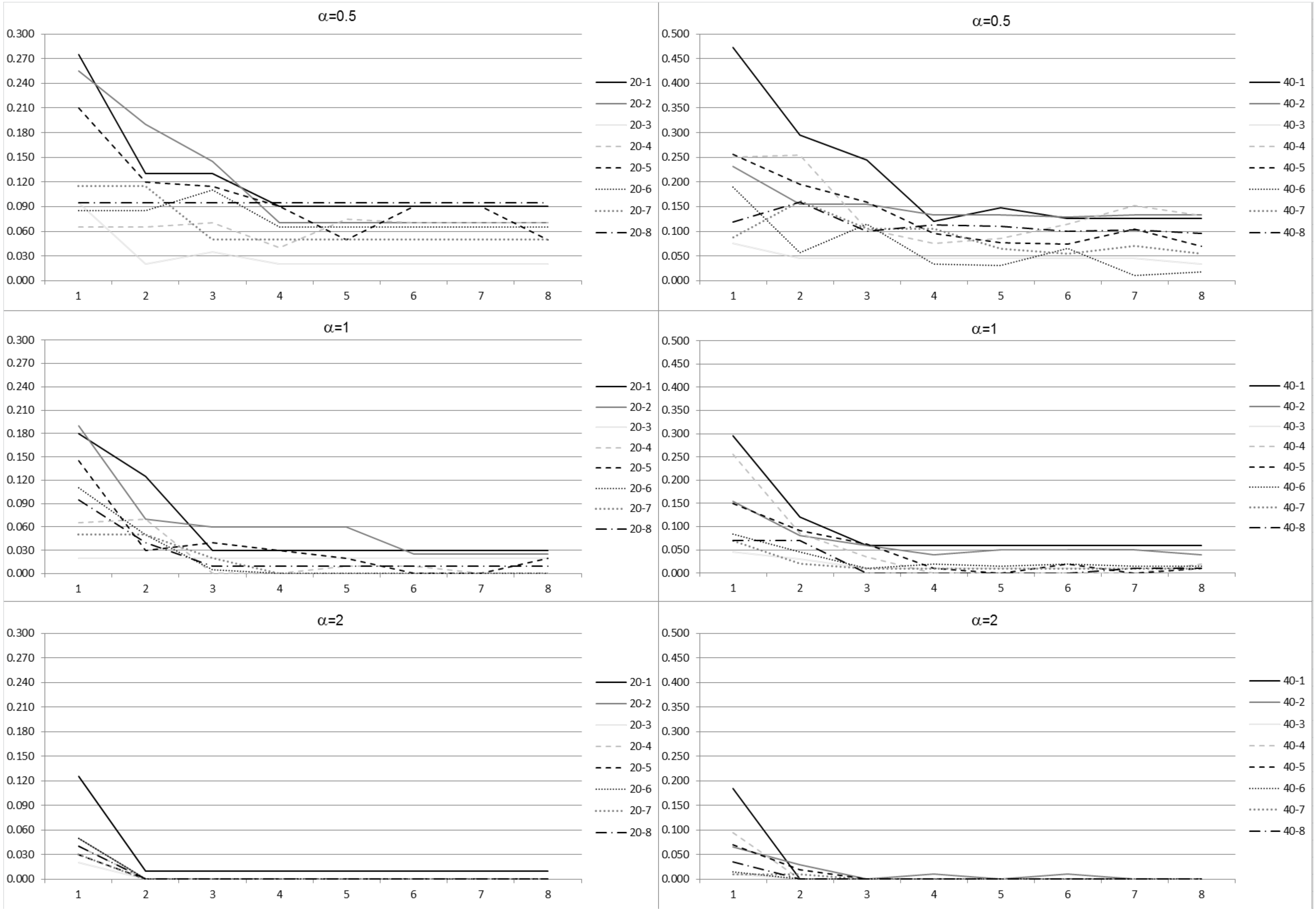

Figure 5: Average number of cancelled patients per block for different values of α

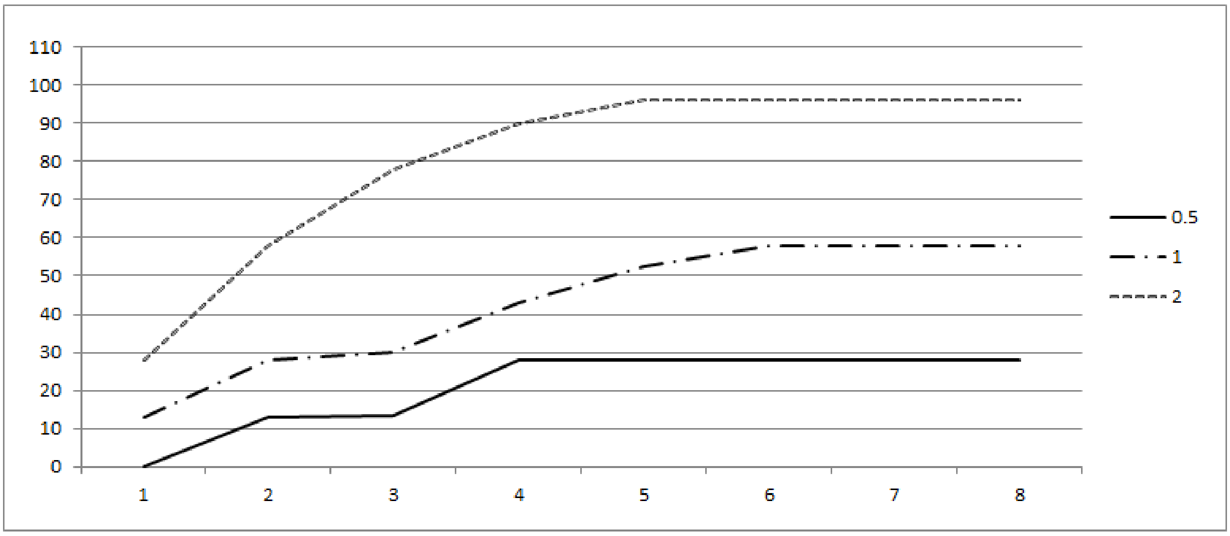

Figure 6: Average per patients delay for different values of α on instance 20-5

3.2.4. Scalability

The last set of tests is run on larger instances, in order to evaluate the scalability of the proposed models. Table 6, 7 and 8 give the computational times and the gaps. The gap is denoted with "ag" when the solver manages to reach the imposed gap of $1.e-3$. The computational time is replaced by "TL" for those instances for which the relative accepted gap is not reached within the two hours time limit. Tables report results for the three and five blocks case. Besides, the average and maximum value is reported for each case.

Concerning the instances with 80 patients (Table 6), the deterministic model is solved to optimality for all the instances. Computational times are negligible for the three blocks case and they increase for the five blocks one. Even for the five blocks case the computational time are about 10 seconds on the average and require more than one minute only in one case.

Concerning the robust model, only five among the instances with 80 patients and three blocks are not solved to optimality. Increasing the value of Γ increases the required computational effort. RM requires very different computational time for different surgery time lists. Instances 80-4 and 80-5 turn out to be the more computationally challenging. Instances solved to optimality require an average CPU time which increases with the increasing value of Γ up to $\Gamma = 6$ and $\Gamma = 7$. Then it slightly decreases for $\Gamma = 7$ and $\Gamma = 8$. This could be due to the fact that for a high value of Γ almost all the patients scheduled require their maximum surgery time. The gap is always below 1.5% for the instances which are not solved to optimality.

Increasing the number of blocks has a significant impact on the computational times. Both the average CPU time and the number of instances not solved within the time limit greatly increase. Besides, 7 instances run out of memory (they are denoted with a "*").

Concerning the robust model, results on 120 patients instances (Table 7) show that the computational time increases when the number of patients increases. Yet, increasing

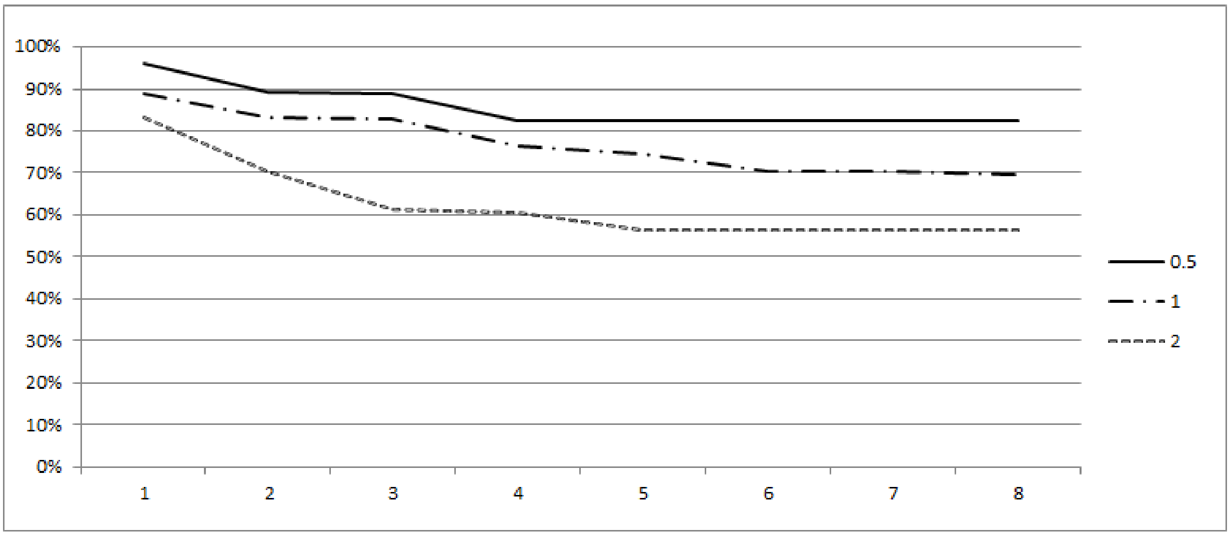

Figure 7: Utilization rate for different values of α on instance 20-5

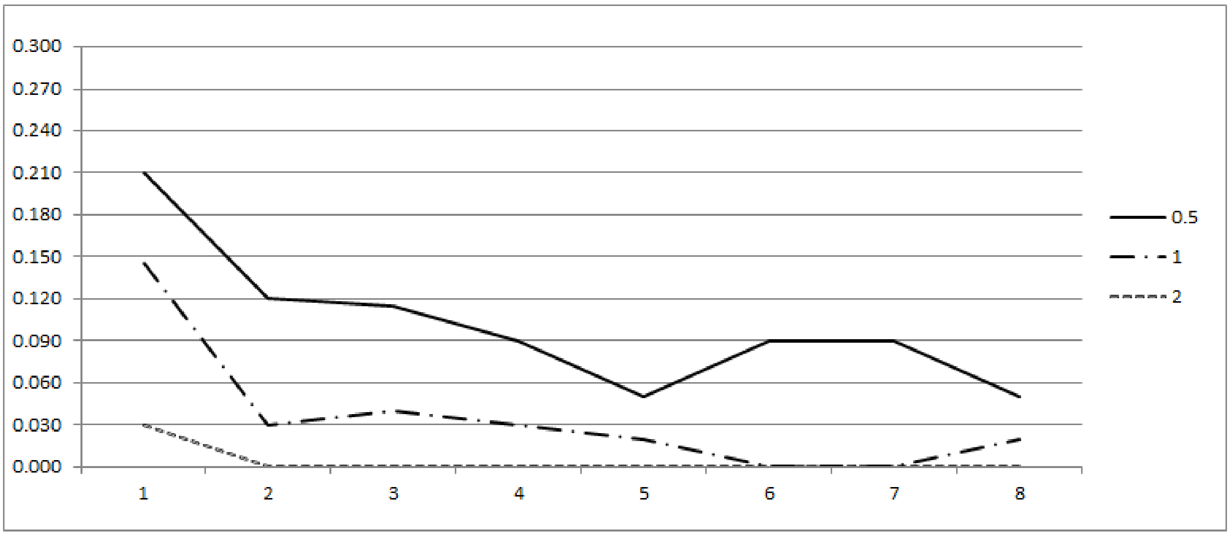

Figure 8: Average number of cancelled patients per block for different values of α on instance 20-5

the number of patients from 80 to 120 seems to have a less significant impact than adding two blocks. All the instances with 120 patients and three blocks are solved to optimality except 7. Such 7 instances show, however, a very low gap. On the contrary, almost all the instances with five blocks cannot be solved to optimality, and three of them run out of memory. However, the gaps are reasonable, being on average between 0.34% and 2.67% and never rising above 6.5%. Deterministic case is not computationally challenging.

A more or less similar behavior is shown for the 140 patients instances (Table 8).

3.3. Overtime impact

We modified SAS to take into account the possibility of considering blocks overtime. We define an amount of overtime allowed for a each block δ. Besides, we assume that at most Δ blocks can exploit overtime in the considered planning horizon. The following variables are added to the models:

- $v_j^k \in \{0, 1\} = 1$ if overtime is assigned to block j in week k
- $o_j^k \geq 0$ amount of overtime in block j of week $k \in K$.

The following constraints are added to the model:

$$\sum_{i \in I} \tilde{t}_i x_{ij}^k \leq \gamma_{jk} + o_j^k \qquad \forall j \in J, \quad \forall k \in K \tag{22}$$

$$o_j^k \leq \delta v_j^k \qquad \forall j \in J, k \in K \tag{23}$$

$$\sum_{j \in J} \sum_{k \in K} v_j^k \leq \Delta \tag{24}$$

Constraints (22) are the capacity constraints for each block forcing either the total time in block j of week k to be less than or equal to the maximum available time γ_{jk} or variable o_j^k to be strictly positive. They replace constraints (3) if overtime is considered. Constraints (23) and (24) limit, respectively, the amount of overtime for each block j and week k and the resulting number of overtime blocks to be less than the a priori fixed values δ and Δ. From this new model for SAS we derived new versions of DAS and SAS taking into account overtime.

We consider a maximum allowed overtime per block equal to 2 hours ($\delta = 120$), and at most $\Delta = \frac{1}{3}|J|$ blocks allowed to use overtime during the planning horizon.

Overtime always improves the solutions. It allows to increase the number of operated patients or to reduce the number of patients operated after their due date. Tables 9 and 10 describe the impact of allowing overtime in terms of utilization rate and cancelled patients, respectively. Results show that allowing overtime increases significantly the utilization rate for most of the cases, with an improvement of up to 10%. As the utilization rate is computed w.r.t. the 6 hours block length, it may rise up to more than 100% for some instances, when overtime is allowed. On the other hand, introducing overtime does not have a strong impact on the number of cancelled patients, although this value decreases slightly for all instances.

3 blocks	DAS		$\Gamma = 1$		$\Gamma = 2$		$\Gamma = 3$		$\Gamma = 4$		$\Gamma = 5$		$\Gamma = 6$		$\Gamma = 7$		$\Gamma = 8$	
instance	gap %	time	gap %	time	gap %	time	gap %	time	gap %	time	gap %	time	gap %	time	gap %	time	gap %	time
80-1	ag	0.05	ag	1.71	ag	0.66	ag	4.33	ag	2.48	ag	2.03	ag	1.59	ag	1.06	ag	0.76
80-2	ag	0.07	ag	0.44	ag	1.89	ag	81.77	ag	24.11	ag	11.76	ag	9.13	ag	5.12	ag	4.76
80-3	ag	0.02	ag	0.13	ag	0.65	ag	0.77	ag	0.65	ag	0.36	ag	0.95	ag	0.33	ag	0.20
80-4	ag	1.13	ag	1.45	ag	95.68	ag	918.61	1.02	TL	1.05	TL	ag	5711.39	ag	2803.27	ag	505.02
80-5	ag	0.09	ag	10.10	ag	46.41	ag	174.70	ag	4570.56	ag	1705.92	1.35	TL	1.17	TL	1.05	TL
80-6	ag	0.05	ag	0.54	ag	0.08	ag	0.16	ag	0.61	ag	0.52	ag	0.34	ag	0.58	ag	0.17
80-7	ag	0.05	ag	0.26	ag	0.09	ag	1.09	ag	20.59	ag	5.97	0.04	5.16	ag	2.92	ag	2.55
80-8	ag	0.11	ag	0.17	ag	0.43	ag	21.28	ag	19.08	ag	5.51	ag	5.42	ag	1.89	ag	1.60
average	ag	0.20	ag	1.85	ag	18.24	ag	150.34	0.21	1478.21	0.22	1114.95	0.25	1615.20	0.23	1250.34	0.22	962.83
max	ag	1.13	ag	10.10	ag	95.68	ag	918.61	1.02	TL	1.05	TL	1.35	TL	1.17	TL	1.05	TL

5 blocks	DAS		$\Gamma = 1$		$\Gamma = 2$		$\Gamma = 3$		$\Gamma = 4$		$\Gamma = 5$		$\Gamma = 6$		$\Gamma = 7$		$\Gamma = 8$	
instance	gap %	time	gap %	time	gap %	time	gap %	time	gap %	time	gap %	time	gap %	time	gap %	time	gap %	time
80-1	ag	2.11	ag	195.93	ag	980.32	ag	1733.23	ag	883.74	ag	605.31	ag	293.38	ag	246.04	ag	163.64
80-2	ag	5.73	ag	1244.76	0.84	TL	2.21	TL	1.92	TL	1.55	TL	1.09	7147.55	0.80	7179.31	0.65	TL
80-3	ag	0.39	ag	55.76	0.38	7187.49	ag	2051.23	ag	1150.95	ag	167.61	ag	168.70	ag	128.74	ag	98.69
80-4	ag	22.05	0.30	7181.18	2.06	7186.28	6.84	TL	7.80	TL	6.99	TL	6.68	5745.76*	5.87	6634.73*	5.26	TL
80-5	ag	66.58	0.39	TL	0.86	5235.98*	2.25	TL	3.14	TL	4.65	790.13*	5.64	TL	5.21	TL	5.08	992.23*
80-6	ag	1.36	ag	13.56	ag	941.52	0.51	TL	ag	7020.63	ag	3616.01	ag	4491.46	ag	2312.77	ag	813.17
80-7	ag	0.58	ag	170.10	ag	187.39	ag	2859.41	0.47	TL	ag	5597.01	ag	4479.43	ag	3293.56	ag	2664.78
80-8	ag	0.53	ag	4.53	ag	7.25	1.65	TL	1.32	6340.00*	0.90	6881.90*	0.83	TL	0.60	TL	0.49	TL
average	ag	12.42	0.16	2006.67	0.57	3614.22	1.72	5322.70	1.87	5518.19	1.81	4004.13	1.83	4587.68	1.61	4271.28	1.49	3286.89
max	ag	66.58	0.39	TL	2.06	TL	6.84	TL	7.80	TL	6.99	TL	6.68	TL	5.87	TL	5.26	TL

Table 6: Percentage gap and computational time for the 80 patients instances

3 blocks	DAS		$\Gamma = 1$		$\Gamma = 2$		$\Gamma = 3$		$\Gamma = 4$		$\Gamma = 5$		$\Gamma = 6$		$\Gamma = 7$		$\Gamma = 8$	
instance	gap	time	gap	time	gap	time	gap	time	gap	time	gap	time	gap	time	gap	time	gap	time
120-1	ag	0.05	ag	2.12	ag	43.89	ag	358.42	ag	197.04	ag	150.19	ag	99.61	ag	54.76	ag	54.21
120-2	ag	0.31	ag	22.82	ag	5.37	ag	1625.98	ag	211.47	ag	78.50	ag	48.97	ag	31.27	ag	10.21
120-3	ag	0.07	ag	0.75	ag	51.25	ag	83.23	ag	52.47	ag	64.26	ag	44.87	ag	23.04	ag	2.78
120-4	ag	0.45	ag	5.61	ag	14.54	ag	5792.23	0.78	TL	0.81	TL	0.45	TL	ag	5227.71	ag	2449.56
120-5	ag	0.09	ag	12.20	ag	98.08	ag	1290.97	ag	1527.07	0.61	TL	1.14	TL	0.93	TL	0.75	TL
120-6	ag	0.19	ag	11.60	ag	81.35	ag	116.02	ag	178.74	ag	82.43	ag	66.48	ag	50.39	ag	14.90
120-7	ag	0.09	ag	1.77	ag	3.42	ag	72.58	ag	115.49	ag	76.87	ag	66.40	ag	39.22	ag	28.46
120-8	ag	0.15	ag	1.08	ag	1.39	ag	2621.18	ag	2375.38	ag	1643.39	ag	725.07	ag	332.66	ag	65.71
average	ag	0.18	ag	7.24	ag	37.41	ag	1495.08	0.18	1480.65	0.25	2058.85	0.27	1928.32	0.20	1618.33	0.18	1226.67
max	ag	0.45	ag	22.82	ag	98.08	ag	5792.23	0.78	TL	0.81	TL	1.14	TL	0.93	TL	0.75	TL
5 blocks	DAS		$\Gamma = 1$		$\Gamma = 2$		$\Gamma = 3$		$\Gamma = 4$		$\Gamma = 5$		$\Gamma = 6$		$\Gamma = 7$		$\Gamma = 8$	
instance	gap %	time	gap %	time	gap %	time	gap %	time	gap %	time	gap %	time	gap %	time	gap %	time	gap %	time
120-1	ag	0.46	1.16	6409.59*	2.45	TL	4.01	TL	2.19	TL	2.04	TL	1.80	TL	1.68	TL	1.19	TL
120-2	ag	7.64	0.44	TL	0.50	TL	2.05	TL	2.23	TL	1.88	TL	1.55	TL	1.27	TL	0.95	TL
120-3	ag	0.44	ag	227.65	1.43	TL	1.76	TL	1.47	TL	0.97	TL	0.77	TL	0.57	TL	0.50	TL
120-4	ag	55.61	0.52	TL	0.68	TL	5.29	TL	6.21	TL	5.72	TL	4.86	TL	4.38	TL	3.89	TL
120-5	ag	0.34	0.17	TL	0.55	TL	1.85	TL	2.87	TL	3.75	2795.42*	4.19	TL	3.86	5965.84*	3.50	TL
120-6	ag	0.59	ag	145.12	0.73	TL	1.62	TL	1.81	TL	1.53	TL	1.21	TL	1.03	TL	0.86	TL
120-7	ag	0.55	ag	50.56	0.27	TL	1.36	TL	1.88	TL	1.52	TL	1.32	TL	1.15	TL	0.94	TL
120-8	ag	0.75	ag	259.97	ag	2760.24	3.06	TL	2.71	TL	2.43	TL	2.12	TL	1.99	TL	1.88	TL
average	ag	8.30	0.34	3581.94	0.84	6634.15	2.63	TL	2.67	TL	2.48	6638.56	2.23	TL	1.99	7034.86	1.71	TL
max	ag	55.61	1.16	TL	2.45	TL	5.29	TL	6.21	TL	5.72	TL	4.86	TL	4.38	TL	3.89	TL

Table 7: Percentage gap and computational time for the 120 patients instances

3 blocks	DAS		$\Gamma = 1$		$\Gamma = 2$		$\Gamma = 3$		$\Gamma = 4$		$\Gamma = 5$		$\Gamma = 6$		$\Gamma = 7$		$\Gamma = 8$	
instance	gap %	time	gap %	time	gap %	time	gap %	time	gap %	time	gap %	time	gap %	time	gap %	time	gap %	time
140-1	ag	0.07	ag	36.47	ag	20.16	ag	283.55	ag	77.67	ag	35.58	ag	19.58	ag	11.82	ag	8.09
140-2	ag	0.13	ag	15.06	ag	51.87	ag	497.16	ag	460.48	ag	168.72	ag	138.32	ag	55.13	ag	61.78
140-3	ag	0.08	ag	1.15	ag	93.00	ag	16.94	ag	25.44	ag	5.73	ag	5.42	ag	2.60	ag	1.46
140-4	ag	0.32	ag	1.89	ag	67.19	0.61	TL	1.03	TL	0.91	TL	0.65	TL	0.26	TL	ag	1494.30
140-5	ag	0.37	ag	20.56	ag	20.97	ag	634.35	ag	471.84	ag	1664.47	0.54	TL	0.42	TL	0.50	TL
140-6	ag	0.11	ag	0.66	ag	14.16	ag	34.42	ag	14.03	ag	6.64	ag	3.59	ag	1.81	ag	1.15
140-7	ag	0.40	ag	0.87	ag	15.20	ag	26.98	ag	65.36	ag	46.07	ag	27.47	0.02	9.51	ag	3.03
140-10	ag	0.06	ag	3.58	ag	1.82	0.18	TL	ag	6826.82	ag	5540.05	ag	1548.02	ag	290.39	ag	71.50
average	ag	0.19	ag	10.03	ag	35.55	0.17	1983.57	0.22	1891.15	0.20	1831.85	0.22	2014.69	0.15	1843.31	0.15	1103.61
max	ag	0.40	ag	36.47	ag	93.00	0.61	TL	1.03	TL	0.91	TL	0.65	TL	0.42	TL	0.50	TL
5 blocks	**DAS**		**$\Gamma = 1$**		**$\Gamma = 2$**		**$\Gamma = 3$**		**$\Gamma = 4$**		**$\Gamma = 5$**		**$\Gamma = 6$**		**$\Gamma = 7$**		**$\Gamma = 8$**	
instance	gap %	time	gap %	time	gap %	time	gap %	time	gap %	time	gap %	time	gap %	time	gap %	time	gap %	time
140-1	ag	0.27	0.78	TL	1.18	TL	2.27	TL	1.08	TL	0.76	TL	0.73	TL	0.64	TL	0.47	TL
140-2	ag	99.08	0.31	TL	0.86	TL	2.13	TL	1.95	TL	1.59	TL	1.21	TL	0.90	TL	0.72	TL
140-3	ag	0.56	ag	60.04	0.98	TL	0.93	TL	0.39	TL	ag	5152.33	ag	1119.23	ag	671.90	ag	166.43
140-4	ag	3.09	ag	34.44	1.21	TL	4.82	TL	5.33	TL	5.01	TL	4.52	TL	3.97	TL	3.64	TL
140-5	0.12	3272.88*	0.24	TL	0.81	TL	1.72	TL	2.51	TL	2.83	TL	3.33	TL	3.11	TL	2.80	TL
140-6	ag	1.26	0.22	TL	0.66	TL	1.03	TL	1.18	TL	0.78	TL	0.69	TL	0.30	TL	0.34	TL
140-7	ag	0.30	ag	148.32	0.41	TL	1.18	TL	1.43	7184.21	1.13	TL	0.98	TL	0.81	TL	0.73	TL
140-10	ag	0.21	ag	6158.32	0.52	TL	2.70	TL	2.45	TL	2.15	TL	1.91	TL	1.70	TL	1.54	TL
average	ag	422.21	0.24	4393.92	0.83	TL	2.10	TL	2.04	7187.16	1.79	6933.17	1.68	6429.03	1.44	6373.13	1.29	6309.93
max	0.12	3272.88	0.78	TL	1.21	TL	4.82	TL	5.33	TL	5.01	TL	4.52	TL	3.97	TL	3.64%	TL

Table 8: Percentage gap and computational time for the 140 patients instances

Instance	DAS	Γ =1	Γ =2	Γ = 3	Γ = 4	Γ = 5	Γ = 6	Γ = 7	Γ = 8
20-1	118.62	90.62	68.71	61.81	-	-	-	-	-
20-2	120.31	102.64	92.31	81.83	82.64	82.09	-	-	-
20-3	119.97	103.03	88.04	83.04	83.30	-	-	-	-
20-4	120.12	106.79	94.00	83.03	78.23	73.57	-	-	-
20-5	116.47	103.44	102.27	94.52	87.54	88.05	88.15	81.98	82.07
20-6	117.64	98.73	96.89	88.73	86.53	-	-	-	-
20-7	115.37	101.53	82.78	90.52	88.44	-	-	-	-
20-8	115.40	105.28	91.47	78.03	79.20	70.95	-	-	-
40-1	119.42	83.51	76.02	57.78	58.30	58.30	57.67	58.30	-
40-2	114.10	97.49	90.46	83.65	79.85	79.28	79.28	79.21	78.96
40-3	112.55	95.16	84.82	80.47	76.64	77.04	77.04	76.64	-
40-4	113.29	103.29	94.70	79.28	73.23	73.50	73.23	73.50	-
40-5	114.21	106.03	98.84	93.43	87.64	83.08	83.29	79.50	78.93
40-6	114.06	101.87	92.24	87.51	85.74	83.29	83.25	83.49	82.97
40-7	106.27	87.03	81.10	75.00	-	-	-	-	-
40-8	109.93	98.18	83.82	75.91	74.72	70.32	70.44	70.42	70.36

Table 9: Utilization rate in percentage if overtime is allowed

Instance	DAS	Γ =1	Γ =2	Γ = 3	Γ = 4	Γ = 5	Γ = 6	Γ = 7	Γ = 8
20-1	0.37	0.25	0.11	0.06	-	-	-	-	-
20-2	0.32	0.19	0.05	0.07	0.03	-	-	-	-
20-3	0.17	0.01	0.01	0.05	0.00	-	-	-	-
20-4	0.28	0.16	0.06	0.02	0.02	0.00	-	-	-
20-5	0.27	0.05	0.07	0.02	0.03	0.03	0.03	0.02	0.00
20-6	0.20	0.07	0.02	-	-	-	-	-	-
20-7	0.26	0.10	0.05	0.03	0.01	-	-	-	-
20-8	0.23	0.15	0.07	0.07	0.03	0.01	-	-	-
40-1	0.20	0.13	0.05	0.07	0.01	0.00	0.00	0.02	0.01
40-2	0.36	0.14	0.08	0.04	0.02	0.04	0.05	0.05	0.07
40-3	0.15	0.05	0.04	0.06	0.01	0.00	0.00	0.01	-
40-4	0.34	0.25	0.12	0.05	0.02	0.01	0.02	0.01	-
40-5	0.30	0.15	0.09	0.05	0.04	0.02	0.02	0.03	0.01
40-6	0.27	0.10	0.05	0.04	0.03	0.03	0.01	0.02	0.01
40-7	0.24	0.09	0.05	0.03	-	-	-	-	-
40-8	0.59	0.24	0.19	0.07	0.07	0.07	0.08	0.07	-

Table 10: Average number of cancelled patients per block, if overtime is allowed

4. Conclusions and further developments

In this paper a robust optimization approach with the Advance Scheduling problem in which surgery times are uncertain parameters is presented. Waiting time, urgency and due date of patients are considered. The goal of the problem is to minimize the penalty associated to waiting time and tardiness of patients.

The proposed models have been tested on a set of real life based instances. The impact of different levels of required robustness is analysed. Besides, we tested the obtained solutions in terms of utilization rate and number of cancelled patients on a set of randomly generated realistic scenarios assuming lognormal distributions for surgery duration. Different choices for the maximum deviation of surgery times are evaluated, as well. Tests on larger instances are run to evaluate the models scalability. Finally, the possibility of allowing overtime is considered, and its impact is evaluated.

Results show that the proposed robust models can be used, as the required computational time is compatible with the weekly schedule. The required computational time are reasonable varying the number patients to be scheduled. Even for larger instances, with up to 140 patients, the gaps are limited. Furthermore, the obtained robust solutions behave well when tested on the scenarios: in fact they improve the number of cancelled patients upon the non robust solution. Although the robust solutions may produce an utilization rate below the 100%, nevertheless by properly tuning the value of parameters, i.e. the level of robustness required, or the degree of risk accepted, a good tradeoff between hospital productivity and quality of service provided to patients can be achieved.

Future work will be devoted to study the impact of different objective function, as well as an to develop an online approach which re-assigns the patients to be rescheduled and deals with the emergency cases.

References

[1] B. Cardoen, E. Demeulemeester, J. Beliën, Operating room planning and scheduling: A literature review, European Journal of Operational Research 201 (2010) 921–932.

[2] F. Guerriero, R. Guido, Operational research in the management of the operating theatre: a survey, Health Care Management Science 14 (2011) 89–114.

[3] B. Cardoen, E. Demeulemeester, J. Beliën, Operating room planning and scheduling problems: A classification scheme, International Journal of Health Management and Information 1 (1) (2010) 71–83.

[4] J. van Oostrum, E. Bredenhoff, E. Hans, Suitability and managerial implications of a master surgical scheduling approach, Annals of Operations Research 178 (1) (2010) 91–104.

[5] V. Augusto, X. Xie, V. Perdomo, Operating theatre scheduling with patient recovery in both operating rooms and recovery beds, Computer & Industrial Engineering 58 (2) (2010) 231–238.

[6] H. Fei, N. Meskens, C. Chu, A planning and scheduling problem for an operating theatre using an open scheduling strategy, Computer & Industrial Engineering 58 (2) (2010) 221–230.

[7] A. Testi, E. Tànfani, G. Torre, A three-phase approach for operating theatre schedules, Health Care Management Science 10 (2007) 163–172.

[8] J. Magerlein, J. Martin, Surgical demand scheduling: A review, Health Services Research 13 (1978) 418–433.

[9] J. Blake, M. Carter, Surgical process scheduling: a structured review, Journal of the Society for Health Systems 5 (3) (1997) 17–30.
[10] V. Augusto, X. Xie, V. Perdomo, Operating theatre scheduling using lagrangian relaxation, European Journal of Industrial Engineering 2 (2) (2008) 172–189.
[11] B. Cardoen, E. Demeulemeester, J. Beliën, Sequencing surgical cases in a day-care environment: An exact branch-and-price approach, Computers & Operations Research 36 (9) (2009) 2660–2669.
[12] H. Fei, C. Chu, N. Meskens, A. Artiba, Solving surgical cases assignment problem by a branch-and-price approach, International Journal of Production Economics 112 (2008) 96–108.
[13] I. Marques, M. Captivo, M. Pato, An integer programming approach to elective surgery scheduling, OR Spectrum 34 (2012) 407–427.
[14] Y. Liu, C. Chu, K. Wang, A new heuristic algorithm for the operating room scheduling problem, Computers & Industrial Engineering 61 (2011) 865–871.
[15] E. Tànfani, A. Testi, A pre-assignment heuristic algorithm for the master surgical schedule problem (mssp), Annals of Operations Research 178 (1) (2010) 105–119.
[16] W. Herring, J. Herrmann, The single-day surgery scheduling problem: sequential decision-making and threshold-based heuristics, OR Spectrum 34 (2012) 429–459.
[17] C. Rizk, J. Arnaout, Aco for the surgical cases assignment problem, Journal of Medical Systems 36 (2012) 1191–1199.
[18] W. Herring, J. Herrmann, Local search for the surgery admission planning problem, Journal of Heuristics 17 (2011) 389–414.
[19] D. Tyler, C. Pasquariello, C. Chen, Determining optimum operating room utilization, Anesthesia and Analgesia 96 (4) (2003) 1114–1121.
[20] S. Choi, W. Wilhelm, An analysis of sequencing surgeries with durations that follow the lognormal, gamma, or normal distribution, IIE Transactions on Healthcare Systems Engineering 2 (2012) 156–171.
[21] B. Denton, J. Viapiano, A. Vogl, Optimization of surgery sequencing and scheduling decisions under uncertainty, Health Care Management Science 10 (2007) 13–24.
[22] D. Min, Y. Yih, Scheduling elective surgery under uncertainty and downstream capacity constraints, European Journal of Operational Research 206 (2010) 642–652.
[23] J. van Oostrum, M. van Houdenhoven, J. Hurink, E. Hans, G. Wullink, G. Kazemier, A master surgical scheduling approach for cyclic scheduling in operating room departments, OR Spectrum 30 (2008) 355–374.
[24] B. Denton, J. Miller, H. Balasubramanian, T. Huschka, Optimal allocation of surgery blocks to operating rooms under uncertainty, Operations Research 58 (2010) 802–816.
[25] E. Hans, G. Wullink, M. van Houdenhoven, G. Kamezier, Robust surgery loading, European Journal of Operational Research 185 (2008) 1038–1050.
[26] E. Tànfani, A. Testi, R. Alvarez, Operating room planning considering stochastic surgery durations, International Journal of Health Management and Information 1 (2) (2010) 167–183.
[27] M. Persson, J. Persson, Analysing management policies for operating room planning using simulation, Health Care Management Science 13 (2010) 182–191.
[28] B. Sobolev, V. Sanchez, C. Vasilakis, Systematic review of the use of computer simulation modeling of patient flow in surgical care, Journal of Medical Systems 35 (2011) 1–16.
[29] C. Banditori, P. Cappanera, F. Visintin, A combined optimizationsimulation approach to the master surgical scheduling problem, Journal of Management Mathematics 24 (2013) 155–187.
[30] M. Lamiri, X. Xie, A. Dolgui, F. Grimaud, A stochastic model for operating room planning with elective and emergency demand for surgery, Journal of Operational Research Society 185 (2008) 1026–1037.
[31] B. Addis, G. Carello, E. Tànfani, A robust optimization approach for the operating room planning problem with uncertain surgery durations, in: International Conference on Health Care Systems Engineering, 2013.
[32] D. Bertsimas, M. Sim, The price of robustness, Operations Research 52 (1) (2004) 35–53.

[33] R. Valente, A. Testi, E. Tànfani, M. Fato, I. Porro, M. Santo, G. Santori, G. Torre, G. Ansaldo, A model to prioritize access to elective surgery on the base of clinical urgency and waiting time, BMC, Health Services ResearchAnnals of Operations Research 9 (1).
[34] W. Hancock, P. Walter, R. More, N. Glick, Operating room scheduling data base analysis for scheduling, Journal of Medical Systems 12 (6) (1988) 397–409.
[35] J. May, D. Strum, L. Vargas, Fitting the lognormal distribution to surgical procedure times, Decision Sciences 31 (1) (2000) 129–148.
[36] D. Strum, J. May, L. Vargas, Modeling the uncertainty of surgical procedure times: Comparison of lognormal and normal models, Anesthesiology 92 (4) (2000) 1160–1167.
[37] F. Dexter, J. Ledolter, Bayesian prediction bounds and comparisons of operating room times even for procedures with few or no historic data, Anesthesiology 103 (2005) 1259–1267.